The National Park Service
Alaska Region
Inventory & Monitoring Program

Data Management Plan
Central Alaska Network

National Park Service
201 1st Avenue
Fairbanks, AK 99701

October 2005

File Name: CAKN_DataMgmtPlan_09162005.doc

Recommended Citation:
Wilder, D.,editor, 2005. Data Management Plan, Central Alaska Network. National Park Service. Inventory and Monitoring Program, Fairbanks, AK. 86 pg, 12 appendices.

Topic(s):
Data management, Planning, Administrative, Interdisciplinary

Theme Keywords:
Reports, protocols, guidelines, plans, data

Placename Keywords:
Alaska, Central Alaska Network, Denali, Yukon-Charley Rivers, Wrangell-St. Elias

Acronyms:

AKRO	Alaska Regional Office
ARLIS	Alaska Resource Library and Information Services
DMP	Data Management Plan
GIS	Geographic Information System
I&M	Inventory & Monitoring (Program)
ITIS	Integrated Taxonomic Information System
NPS	National Park Service
PDF	Adobe Portable Document Format
CAKN	Central Alaska Network
YUGA	Yukon-Charley Rivers, Gates of the Arctic main office
LAN	Local Area Network
ERD	Entity Relationship Diagram (for databases)

Initial Distribution:
Website: *www.nature.nps.gov/im/units/cakn/DataMgt.htm*

Table of Contents

List of Figures

List of Tables

Executive Summary

Chapter I: Introduction

- As a critical component of the CAKN Monitoring Program, the Data Management Plan (DMP) aims to:
    - Outline the long-term goals of a comprehensive data management strategy for the CAKN I&M Program
    - Associate data management goals with the long-term goals of the network and service-wide I&M program
    - Outline the procedures and work practices that support effective data management
    - Guide current and future staff of the CAKN to ensure that sound data management practices are followed
    - Guide the enhancement of legacy data to match formats and standards put forth in this plan
    - Encourage effective data management practices as an integral part of project management so all data are made available and usable for park management decisions now and into the future
    - Optimize and promote interagency sharing and development of data, software applications and analyses
    - Establish roles and responsibilities of program staff for managing data
    - Identify necessary elements for a functional data management program and describe any anticipated changes to those elements
    - Establish an organization schema for Program data and information so that they are retrievable by staff, cooperators and the public
    - Establish basic quality control standards
    - Establish standards for data, data distribution and data archiving to ensure the long-term integrity of data, associated metadata and any supporting information

- Specific data and information the CAKN Program deems necessary to meet objectives includes:
    - Core variable data measured in the field
    - Data derived via vital sign protocols from core variable data
    - Spatial data files
    - Photographs (field and aerial)
    - Laboratory data
    - “Data” or “Technical” Reports including protocols
    - Administrative Reports
    - Field data sheets, books
    - Selected external and legacy data and datasets

- The CAKN data management strategy draws from national-level I&M guidelines and formalizes them as policy at the network level. More detailed data management strategies are documented in standard operating procedures specific to a given data collection effort.

- As a plan, the DMP will likely undergo significant revision as CAKN operations are implemented. Informal review and revision will be on-going to meet the

changing needs of the CAKN I&M Program. The latest version of the plan is available on the CAKN website at *www.nature.nps.gov/im/units/cakn/DataMgt.htm*.

Chapter II: Process and Workflow

- Core data management operations are conducted within the basic process and workflow of a data-generating project. Projects are divided into the following stages:
    1. Project Initiation
    2. Planning and Approval
    3. Design and Testing
    4. Implementation
    5. Product Delivery
    6. Product Integration
    7. Close Out and Evaluation

- The CAKN will use a Project Tracking database to track basic project parameters.

Chapter III: Management Roles and Responsibilities

- For the CAKN Program to work effectively, everyone within the Network will have stewardship responsibilities in the production, analysis, management, and/or end use of the data.

- Each Vital Sign monitoring protocol and any project study plan contains specific instructions for assignments and tasks that nest within the overall framework of the Data Management Plan. Individuals who carry out monitoring protocols and inventory study plans are responsible for reading and understanding these instructional guidelines.

- The fundamental role of the Network data manager is to understand and determine program and project requirements, to create and maintain data management infrastructure and standards, and to communicate and work with all responsible individuals.

- The Project Leader (PL) oversees and directs operations, including data management requirements, for one or more Network projects. The PL maintains communication with project staff, Network Data Manager, and resource specialist regarding data management. Project leaders are responsible for designating an alternate leader who is capable of maintaining project operations in his or her absence.

- The Network Coordinator interfaces with project leaders to ensure that timelines for data entry, validation, verification, summarization/analysis and reporting are met. Additionally, the Network Coordinator must review and approve proposed changes to project protocols prior to implementation.

Chapter IV: Infrastructure

- The CAKN monitoring program relies heavily on park, regional and national information technology (IT) personnel and resources to maintain the computer resource infrastructure.

- The CAKN data management "system" is currently (10/1/2005) in development. The system design presented here represents current plans which may be significantly altered by the point of full implementation.

- The CAKN will establish and operate a server database (MS SQL Server) housing electronic data and information managed by the program. This relational database will allow staff to browse, evaluate, export, analyze and integrate vital sign monitoring data and information for research, management and reporting purposes. Development of this system is planned in three stages culminating in conversion of all data to GeoDatabase format for serving via the ESRI Spatial Database Engine. GIS and web mapping applications will be implemented to enhance data access.

- The Alaska Regional Office provides the following that the CAKN will utilize to meet its goals:
    - The Wide Area Network file server for general file exchange and storage
    - GIS and related tabular data accessible via custom applications distributed to the parks as well as the Alaska GIS Data Clearing House (*www.nps.gov/akso/gis*).
    - File server to provide offsite storage for all CAKN data

Chapter V: Acquisition and Processing

- Data designs for all CAKN monitoring data and CAKN-initiated projects will conform to the I&M Natural Resource Template Standards and confine, where possible, data values to specific ranges.

- Data discovery is an on-going process requiring regular data searches and visits to Network parks in order to ensure that the CAKN I&M Program maintains as much relevant material pertaining to the parks as possible. Data acquired by the CAKN will be developed, if necessary, as digital datasets conforming to NPS and I&M database standards.

- Upon the completion of any field work, a summary of what was collected will be entered in a project tracking database.

- Acquisition of data sourced outside the NPS will be addressed in the appropriate Vital Sign protocol.

- Data and information for the CAKN may be organized into five categories: Monitoring, Inventory, Prototype (DENA LTEM), Short-term (e.g. pilot projects) and Legacy (primarily "mined" data).

- The CAKN intends to scan hardcopy references and materials, saving them as .pdf files, in order to create a digital library.

- Each CAKN project uses a working database developed in MS Access to perform initial data processing (entry, verification, validation, metadata generation) after acquisition.

- As part of data collection and entry duties, the project leader and data manager will ensure that data from laboratories match project data designs and formats. They will also build into the project data design a mechanism for relating instances of field work (data collection events), raw data, and laboratory results. This will be done via sample and specimen labels that can be matched with field work event identification numbers.

- Significant changes to the protocols dictating data acquisition methods must be approved by the project leader, network coordinator and the data manager. The network coordinator must evaluate the proposed changes and determine if additional peer review is required before approving.

Chapter VI: Assurance (QA) and Quality Control (QC)

- Data management for the CAKN Inventory and Monitoring program must ensure that our projects produce and maintain data of the highest possible quality. The CAKN will develop a comprehensive set of SOPs and tools for quality assurance and control in field procedures, data entry/validation/verification and data use (browsing, sub setting, downloading, analysis, etc.).

- NPS Director's Order #11B specifies standards that apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

- Project subjects and goals will drive data quality needs and control the kinds of analysis and summarization that may be defensibly applied.

- Laboratories that will be entering analysis results for a given vital sign will be supplied with a copy of the database application so that data may be entered in the manner and format matching that of the rest of the data for a given monitoring parameter. While most professional laboratories exercise their own QA/QC procedures, results received by a project leader are subject to the same QA/QC measures exacted on other project data.

- As a standard part of database design, the CAKN will build into database tables, fields that track at the record-level who entered the data, precise entry time and the protocol version under which the data were collected.

- Data Collection: At a minimum, data will be collected on formatted, project-specific data sheets that reflect the overall data design for the project and maximize limitations on values that may be recorded for different parameters. Sheets will be designed to minimize the amount of writing necessary to effectively record observations.

- Data Entry: Data will be entered as soon as reasonably possible after collection by someone familiar with data collection. Data will be entered into pre-designed database applications that resemble field sheets and maximize error control. Data will not be entered into spread sheets. The CAKN will maximize the use of database programming to control data entry. To the extent possible, data entry will be automated.

- Data Verification: Data verification is carried out by staff sanctioned by the project leader who are ideally familiar with data collection and entry. One hundred percent of records will be verified against original source data. Ten percent of records will be reviewed after initial verification by the project manager. If errors are found, the entire data set verified again. A record of each dataset's verification process including number of verification iterations and results will be prepared by the project leader as part of formal metadata generation.

- Data Validation: Corrections or deletions as a result of data validation require notations in the original paper field records and in any copies made for data entry about how and why the data were changed. The CAKN will maximize the use of automated routines and/or data summary/visualization such as histograms, line plots, and basic statistics to reveal possible logic and range errors.

- The data manager will conduct periodic "spot checks" of random CAKN monitoring projects to ensure compliance with data management plan and project protocol QA/QC procedures.

Chapter VII: Documentation

- Formal, standard metadata that complies with federal and NPS standards serve as the principal documentation for CAKN data.

- The CAKN data manager will establish a standard operating procedure for metadata generation and maintenance to be used by all CAKN projects.

- In general, a single metadata document will apply to both raw and certified versions of the data. Metadata records will be stored with both hard copy and digital archive data.

- Generally, metadata will be created by the project leader, with assistance from the data manager.

- Metadata will be served to the internet via the CAKN primary data server out of Fairbanks as well as the national server in Ft. Collins (NR-GIS metadata server). The CAKN data manager ensures metadata are up-to-date on all servers.

Chapter VIII: Analysis and Reporting

- The CAKN strategy towards data analysis and reporting rests upon providing sufficient funding for these activities so that they occur promptly—that is, to report on the previous phenological year (Oct-Sept) by the following March.

- The CAKN will also focus on producing an annual integrated "State of the Parks" report that effectively communicates the changes and trends observed in each Vital Sign to our primary audience—the natural resource managers of each park.

- Each CAKN project protocol will include a standard operating procedure addressing data analysis. Development of these will:
    1. Use straightforward, minimally structured sampling designs.
    2. Work closely with statisticians in developing and implementing change detection analyses.

3. Provide adequate support to project leaders for data analysis (including hiring of staff).

- In general, analysis steps are:
  1. Summarization
  2. Outlier detection
  3. Change over time
  4. Vital Sign relationships
  5. Time series analysis

- The main audience for monitoring data is the resource managers of each network park and other managers in the National Park Service system, who will use the information to assist with their management decisions.

- The CAKN vision for reporting includes the following central themes: (1) We will prepare monitoring reports that are understandable and useful to our primary audience: park resource managers, (2) We will prepare reports promptly, and (3) All reports will be readily available.

Chapter IX: Dissemination

- Data management within the CAKN I&M Program aims to ensure that
  - Data are easily discoverable and obtainable
  - No data that have not been subjected to full quality control are released
  - Distributed data are accompanied by complete metadata which clearly establishes the data as a product of the NPS I&M Program
  - Sensitive data are identified and protected from unauthorized access and inappropriate use
  - A complete record of data distribution/dissemination is maintained

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.

- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.

- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

- Network staff will classify sensitive data on a case-by-case, project-by-project basis.

- All references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

- Repositories for CAKN data, products and selected project items include:
  - CAKN Primary Data Server
  - NPS Natural Resources GIS Metadata and Data Store (NR-GIS)

  - NPS on-line bibliographic database (NatureBib)
  - NPS on-line document server (NPS Focus)
  - Alaska Resources Library and Information Services (ARLIS)
  - EPA STORET (on-line water quality database)
  - University of Alaska, Fairbanks Museum (primarily for plant specimen)

- In general, data will be available upon completion of analysis and reporting. Data for which analysis and reporting has not been completed but are otherwise certified (verified and validated) will be released no later than one year after certification.

Chapter X: Maintenance, Storage and Archiving

- CAKN data maintenance, storage and archiving procedures aim to ensure that data and related documents and materials (digital and physical) are
  - Kept up-to-date with regards to content and format such that the data are easily accessed and their heritage and quality easily learned.
  - Physically secure against environmental hazards, catastrophe, and human malice
  - Archived in a manner that expedites recovery if needed

- Most maintenance activity will involve active monitoring datasets however finalized CAKN-project data will also be maintained along with active data in a common relational database system.

- Primary digital data maintenance will be performed on the main CAKN server in Fairbanks.

- A catalogue of the data and information on the CAKN server will be maintained on the CAKN website and reflect changes or updates to datasets.

- All digital files on the primary CAKN server will be backed up daily in Fairbanks. These data will also be backed up via the regional network to a server in Anchorage at least weekly. All digital files will be restorable from backup sources in either Fairbanks or Anchorage. Both of these backup sources will be read-only and accessible for data restoration purposes only.

- Raw, certified and analyzed data (data products) will be archived and a common metadata file will be associated with each.

- Datasets that are considered complete and inactive will be saved on the primary CAKN server in both ASCII and native formats. Data are considered complete and inactive when accompanying metadata, as generated by the project leader or other authorized personnel, list the project status as “complete”.

- Hard copy project materials (including project reports) will be archived according to NPS standards, policy and procedure.

- Photographs from each CAKN project will be entered into a database where attributes such as electronic file name, keywords, project, description, photographer, date and location will be catalogued. Digital photo management

will generally follow guidelines established in a draft management strategy prepared by the Alaska Southeast and Southwest I&M Networks.